CMLS, Cell. Mol. Life Sci. 60 (2003) 277–287
1420-682X/03/020277-11

CMLS Cellular and Molecular Life Sciences

Cyanovirin-N: a sugar-binding antiviral protein with a new twist

I. Botos and A. Wlodawer *

National Cancer Institute, MCL, Bldg. 536, Rm. 5, Frederick, Maryland 21702-1201 (USA), Fax +1 301 846 6128, e-mail: wlodawer@ncifcrf.gov

Abstract. Cyanovirin-N (CV-N), an 11-kDa protein from the cyanobacterium *Nostoc ellipsosporum,* is a highly potent virucidal agent that has generated interest as a lead natural product for the prevention and chemotherapy of human immunodeficiency virus infection. The antiviral activity of CV-N is mediated through specific, high-affinity interactions with the viral surface envelope glycoproteins. A number of structures of wild-type, mutant and sequence-shuffled CV-N have been solved by nuclear magnetic resonance and crystallography, showing that the protein exists as either a quasi-symmetric two-domain monomer or a domain-swapped dimer. Structures of several complexes of CV-N with oligosaccharides help in explaining the unique mode of high-affinity binding of these molecules to both forms of CV-N.

Key words. Cyanovirin-N; anti-HIV activity; 3D domain-swapping; gp120; protein folding; X-ray; NMR.

Introduction

Microorganisms are an excellent source of biologically active compounds that include polyketide and peptide antibiotics, among others [1]. The proteinaceous constituents of natural product extracts represent a largely untapped source of potentially novel, biologically active molecules. Perhaps the most significant reason that proteins from natural product extracts have not been studied more fully is the widely held belief that these bioactive molecules do not have the pharmacological characteristics necessary to become useful drugs [2].

Cyanovirin-N (CV-N) is a novel protein, originally isolated from cultures of the cyanobacterium (blue-green algae) *Nostoc ellipsosporum* [3], that has generated interest as a lead natural product for the chemotherapy of HIV infection [4]. CV-N is an 11-kDa protein with a primary amino acid structure that has no significant homology (less than 20%) to any known protein. CV-N shows a remarkable degree of stability and activity after several freeze-thaw cycles, as well as after treatment with organic solvents (CH_3CN, CH_3OH, DMSO), denaturants (8M GnHCl), detergents (0.5% SDS), 0.5% H_2O_2 or even 15 min boiling [3]. In addition to being isolated from its natural source, CV-N was also expressed from recombinant DNA in *Escherichia coli* [5], and most of the experiments described below used the recombinant protein.

Both natural and recombinant CV-N are highly potent virucidal agents. They irreversibly inactivate diverse T-lymphocyte-tropic, macrophage-tropic and T-lymphocyte- and macrophage-tropic primary isolates of human immunodeficiency virus (HIV)-1, HIV-2, simian immunodeficiency virus (SIV), feline immunodeficiency virus (FIV) and certain other enveloped viruses [3, 6]. CV-N inhibits in vitro fusion of HIV-infected and noninfected cells and cell-to-cell transmission of HIV-1 infection [3]. The antiviral activity of CV-N is mediated through specific, high-affinity interactions with the viral surface envelope glycoproteins, gp120 [3] and possibly also gp41 [2, 7]. It is thought that binding of CV-N renders these glycoproteins incapable of mediating virus-to-cell or cell-to-cell fusion.

In this review we will discuss the unique structure of CV-N, in particular the transition between monomeric and dimeric state of this protein. Carbohydrate-binding properties of the native protein and of a number of its mutants will be discussed and correlated with the biological properties of this promising candidate for a drug that could prevent HIV infection in a completely novel manner.

* Corresponding author.

Primary structure of CV-N

CV-N shares less than 20% sequence homology with any other known or hypothetical protein. Internally, however, the sequence shows a high degree of duplication, with clear similarity found between part A, comprising residues 1–50, and part B, consisting of residues 51–101 (fig. 1A). A comparison of the amino acid sequences of these chain segments shows that 16 residues (32%) are identical and another 13 residues (26%) show conservative amino acid substitutions. The highest degree of homology is in the loops defined by the disulfides Cys8–Cys22 and Cys58–Cys73, and in the region farthest from these loops (residues 40–49 in A, and 91–100 in B). At this level of sequence conservation, the structures corresponding to these segment chains were expected to be similar, and indeed that turned out to be the case, although in a rather convoluted way (see below). Similar repetition of amino acid sequences has also been observed in many other proteins and is believed to result from gene duplication during evolution [8, 9].

3D structures of CV-N

A number of structures of CV-N have been solved under a variety of conditions involving both solution and crystalline states. Table 1 summarizes available data for three-dimensional (3D) structures of the wild-type and mutant forms of CV-N determined by nuclear magnetic resonance (NMR) [10–14] and by X-ray crystallography [14–16a]. The quality of solution structures of CV-N is very high, since they were solved using double and triple resonance multidimensional heteronuclear NMR with ^{15}N and ^{15}N/^{13}C-labeled protein. In addition, a series of novel restraints were utilized, including ^{1}H and ^{13}C shifts, ^{3}J couplings and residual dipolar couplings [17]. A number of crystal structures of CV-N have also been solved in the last few years, some of them to high resolution (up to 1.5 Å). The initial crystal structure was solved from trigonal crystals (space group $P3_221$) grown at low pH [15], whereas a tetragonal crystal form of CV-N (space group $P4_12_12$) was grown at high pH [14]. It is important to emphasize the conditions under which the crystals were grown since some of the structural features appear to be pH dependent.

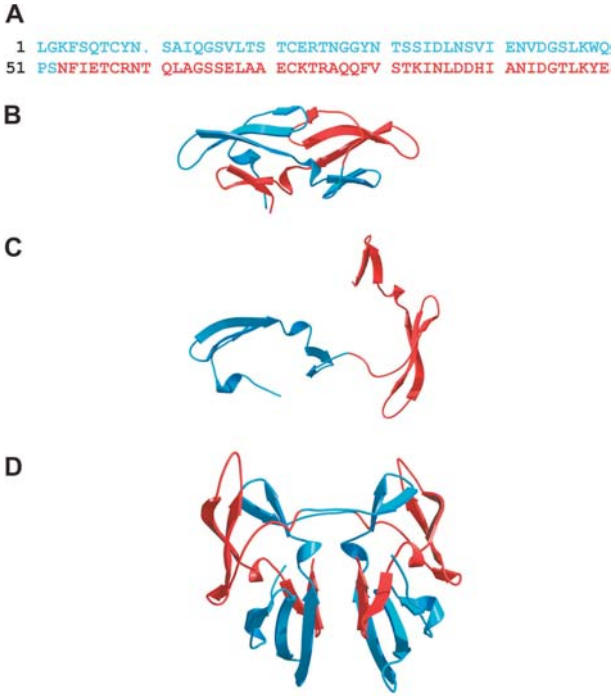

Figure 1. Structural forms of CV-N. (*A*) Sequence alignment of the two CV-N domains, with the color-coded domains mapped to the structure of the (*B*) compact monomer (2ezm), (*C*) domain-swapped extended monomer (part of 3ezm), (*D*) domain-swapped dimer (1l5b). Figures generated with programs Bobscript [56] and Raster3D [57].

The first solution structure of CV-N indicated that the protein was monomeric and consisted of two structurally very similar domains [10]. However, these domains did not correspond to the amino acid sequence of the protein in a linear fashion, since the first domain was made up of residues 1–38 and 90–101, whereas the second domain consisted of residues 39–89. Since the structure of the hairpin 90–101 is almost identical to that of 39–50, the monomer has been formed by intramolecular strand exchange, a process described in detail for serpins [18], although with some significant differences. The situation became even much more complicated when a dimer consisting of two molecules of CV-N was found in the crystal structure [15]. This dimer represents an unusual case of an almost symmetric domain swapping and consists of two quasi-monomers resembling the one described above, with one made up of residues 1–50 (A) and 51′–101′ (B′) and the other of 51–101 (B) and 1′–50′ (A′). Similar dimers were later also seen in some NMR structures [13, 14, 19].

In what follows, we will disregard the strand-exchanged nature of the monomer of CV-N, since this feature is preserved between all known structures of this protein. Domain swapping, however, will be invoked to explain the dimeric state of the molecule, if present. A monomer of CV-N (or a quasi-monomer of a domain-swapped dimer) can be described as a quasi-symmetric molecule with an ellipsoidal shape, ~55 Å in length with a maximum width of ~25 Å (fig. 1B). The two similar domains of the monomer are made up predominantly of β strands and a few short 3_{10} helical turns linked by two intramolecular disulfide bonds (Cys8–Cys22 and Cys58–Cys73). These disulfide bonds are important for the structural stability and anti-HIV activity [9]. Superposition of residues 1–50 and 51–101 results in a root-mean-square deviation of 1.3 Å for all Cα carbons in the trigonal crystal

Table 1. Currently available structures of CV-N.

PDB ID	Method, reference	Resolution (Å)	Ligand	CV-N	Oligomeric form	Space group
2EZN	NMR 10	–	–	wild-type	compact monomer	–
2EZM	NMR 10	–	–	wild-type	compact monomer	–
3EZM	X-ray 15	1.5	–	wild-type	domain-swap dimer	$P3_221$
1J4V	NMR 13	–	–	wild-type	domain-swap dimer	–
1L5B	X-ray 14	2.0	–	wild-type	domain-swap dimer	$P4_12_12$
1L5E	NMR 14	–	–	wild-type	domain-swap dimer	–
1IIY	NMR 11	–	Manα1-2Manα	wild-type	compact monomer	–
1M5M	X-ray 16a	2.5	Man-9	wild-type	domain-swap dimer	$P4_12_12$
1M5J	X-ray 16a	2.4	hexamannose	wild-type	domain-swap dimer	$P4_12_12$
1LOM	X-ray 16	1.72	–	P51S/S52P mutant	domain-swap dimer	$P3_221$
1NO2	NMR 12	–	–	circularly permuted (cpCV-N)	compact monomer	–
–	NMR 38	–	–	sequence-swapped (dsCV-N)	compact monomer	–

structure [15]. A search of the Protein Data Bank using the program DALI [20] did not reveal any proteins with significant structural similarity to CV-N, although it bears distant resemblance to the hyperthermophile DNA-binding protein Sac7d [21] and the SH3 domain of spectrin [22].

Monomers, dimers and domain swapping

As mentioned above, a major difference between the initial solution and crystal structures of CV-N was the 3D domain swapping, observed in the latter case. Due to a change of torsion angles in the hinge region (residues 49–54), domains A and B of CV-N separate into a very elongated entity in which they do not contact each other (fig. 1C). Two such symmetrically related monomers form a dimer in which domain A from one monomer interacts with domain B′ from the other monomer and vice versa, forming two quasi-monomers, AB′ and A′B, linked by their respective hinge regions (fig. 1D). This domain-swapped dimer is quite unique in that half of each molecule takes part in its creation (in most other cases described to date, only a small part of the protein is involved in domain swapping [23]). The quasi-monomers can adopt a range of orientations relative to each other, depending on the conditions of the experiment. A significant change in the orientations of the domains in the low-pH and high-pH structures is illustrated in figure 2.

The reasons behind these large differences in the orientations of the domains have not yet been satisfactorily explained. It has been shown that proline residues are able to influence protein conformation [24] and, when located in a hinge region, may promote arm exchange and domain swapping for oligomerization of monomeric proteins [25]. However, the importance of these proline

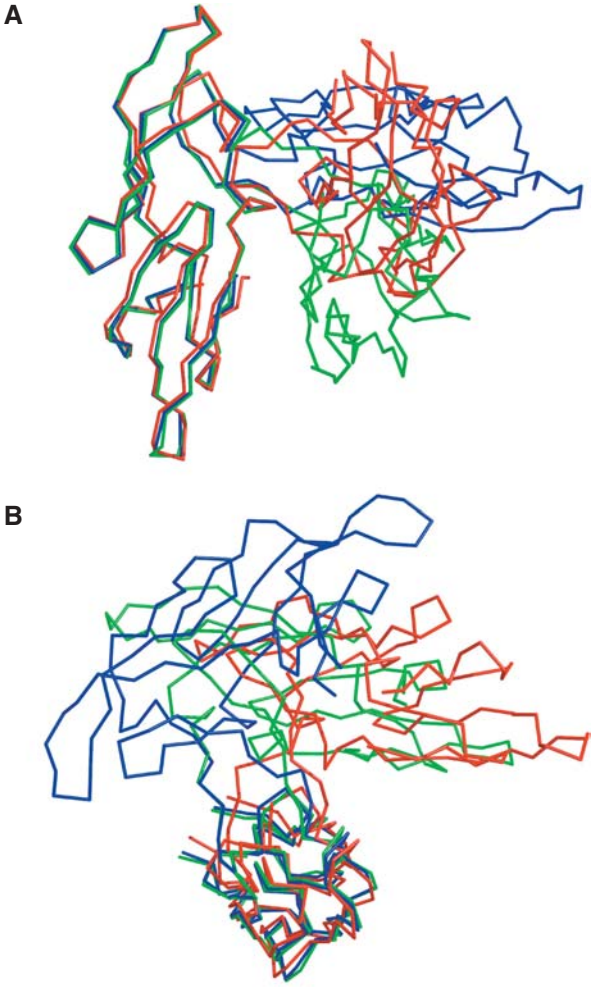

Figure 2. Comparison of pseudo-monomer orientation in CV-N structures at different pH values. (*A*) Side view. (*B*) Top view. The same pseudo-monomer from the following structures was superimposed: low pH crystal structure (blue), high pH crystal structure (green) and pH 6.4 solution structure (red). Structures aligned with program ALIGN [58].

residues was put into a different light by results describing the crystal structure of a CV-N double-mutant P51S/S52P [16]. In the P51S/S52P structure the hinge region adopts a significantly different orientation than in the wild-type domain-swapped dimer, although the relative orientation of quasi-monomers is the same as in the low-pH crystals of wild-type CV-N. However, this mutant adopts the same relative orientation of the quasi-monomers under a wide range of pH conditions, challenging the hypothesis that domain swapping is a result of low pH and further questioning the relative importance of the Pro residue in the hinge region as a determinant of domain orientation.

A surface and volume analysis of some of the current CV-N structures [16] showed a trend of more buried atoms with the decrease of pH, yielding a more compact structure at low pH. At the same time, the solution structures proved to be more compact than the crystal structures. These observations suggest a domain orientation determined by electrostatics of the whole protein rather than a single Pro residue in the hinge region. The relative orientation of the domains may be important, since it might influence the oligosaccharide-binding affinity of CV-N (see below). The dimer of CV-N was shown to be a metastable, kinetically trapped intermediate at neutral pH and room temperature, under which conditions it is stable for months. After incubation for 24 h at 38 °C, the metastable dimer converts into a thermodynamically more stable monomer [14].

A number of ions have been observed in the native and mutant CV-N crystal structures and they appear to have a stabilizing effect [16]. The high-pH tetragonal structure has a well-defined Na ion close to the hinge region, anchored by OE1 of the side chain amide of Gln50 from domain A and the main-chain carbonyl O of Ser38 from domain B. This ion was, however, not observed in the low-pH trigonal structure of the native CV-N and in the P51S/S52P mutant. In the P51S/S52P structure there is a well-defined sulfate near the N-terminus, with a well-conserved position among structures at different pH values.

Oligosaccharide binding to CV-N

Two carbohydrate-binding sites with different affinities have been mapped on the surface of CV-N [26]. These sites are created by structurally equivalent residues belonging to the two domains of monomeric CV-N and bind primarily N-linked high-mannose oligosaccharides, such as those found on the viral envelope of HIV-1 [27]. In monomeric CV-N, the primary carbohydrate-binding site is defined by residues 41–44, 50–56 and 74–78, and the secondary site by residues 1–7, 22–26 and 92–95 [26]. There are four sugar binding sites in a domain-swapped dimer: two primary sites, located close to each other near the hinge region, and two secondary sites on the opposite sides of the dimer, not influenced directly by the conformation of the hinge region. In a monomer, intertwined loops from domains A and B form the sugar-binding sites, whereas in a domain-swapped dimer these loops contain residues from both monomers. In this manner, residues 1–7 and 22–26 from the first monomer and 92′–95′ and 101′ from the second monomer form the secondary site. Composite active sites resulting from domain swapping have been reported in other cases [28–30].

The structure of the primary sugar-binding site of CV-N was first described in solution, in the presence of the disaccharide Manα1-2Manα [11]. The disaccharide binds in a deep pocket in the close proximity of the hinge region

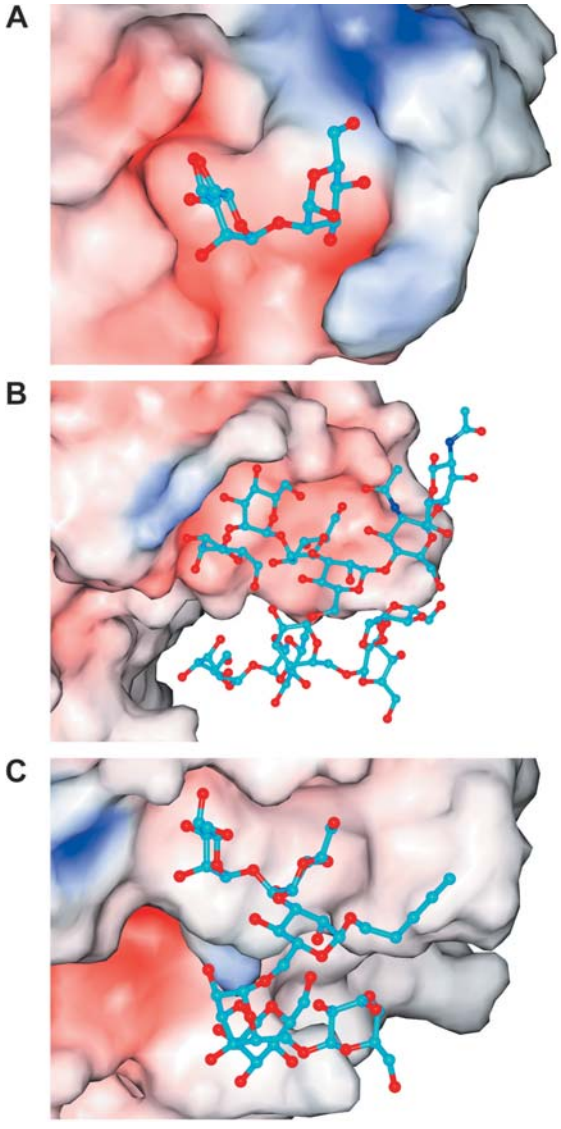

Figure 3. Structure of CV-N complexed to oligosaccharides. (*A*) Manα1-2Manα dimannose; (*B*) Man-9; (*C*) hexamannose. Figures generated with program SPOCK [59].

in a stacked conformation, establishing hydrogen bonds with Glu41, Asp44, Ser52, Glu56, Thr57, Thr75, Arg76 and Gln78 (fig. 3A). No sugar binding in the primary sugar binding site as defined above has been detected in the recently solved crystal structure of a complex of CV-N with an oligosaccharide (fig. 4) [16a]. Instead of a sugar, this site contained a well-defined, tightly bound buffer molecule (CHES) from the crystallization solution. The CHES molecule does not bind into the secondary sugar binding site, not even in the uncomplexed structures. It is unlikely that the CHES prevents oligosaccharide binding, since it does not exhibit measurable binding affinity in solution. It partially obstructs the binding pocket, forming a strong hydrogen-bond anchoring the sulfate oxygen atom O13 to Arg76 NH2 of the protein [16a].

Comparison of available crystal [14, 15] and solution structures [10, 11] of CV-N reveals the changing geometry of the primary sugar binding site upon domain swapping. We can only speculate on the significance, if any, of this shift in the relative orientation of the two domains. Since the primary sugar binding site is in close proximity to the hinge region, the position of the hinge and relative orientation of the domains has a direct impact on the shape of the primary sugar binding site. In the monomer the pocket is intact, accommodating a disaccharide in a stacked conformation [11], whereas in all domain-swapped structures some of the essential protein-oligosaccharide hydrogen bonds cannot be established, leaving only the shape complementarity of oligosaccharide and pocket (fig. 5). Specifically, in the low pH (trigonal) crystal structure, the sugar-binding pocket is partially changed compared with the one observed in the compact monomeric structure (fig. 5A). The OG moiety of Ser52 is displaced from its optimal position in which it provides a hydrogen-bond to an oxygen atom from a mannose ring, disrupting one of the important hydrogen bonds and making the binding of any saccharide sterically unfavorable (fig. 5B). This is the case in all trigonal crystal structures crystallized at pH range 4.6–8.5, which adopt the same

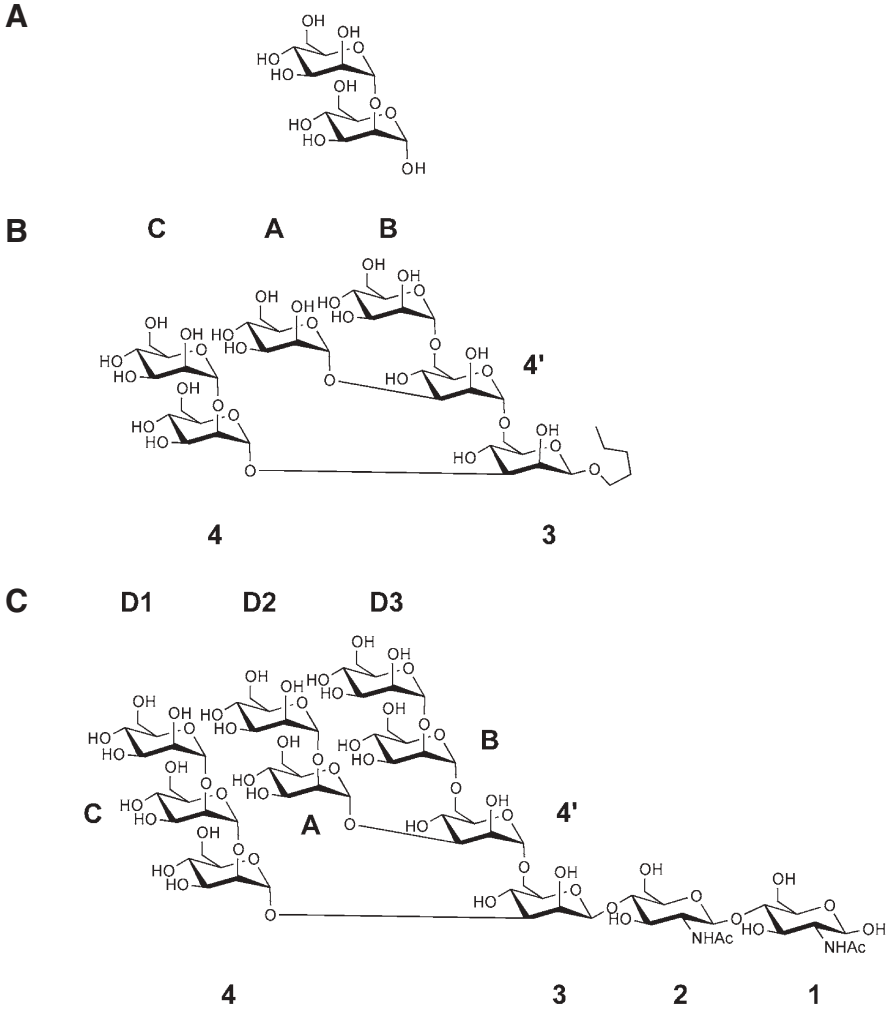

Figure 4. Chemical structure of oligosaccharides. (*A*) Manα1-2Manα dimannose; (*B*) hexamannose; (*C*) Man-9.

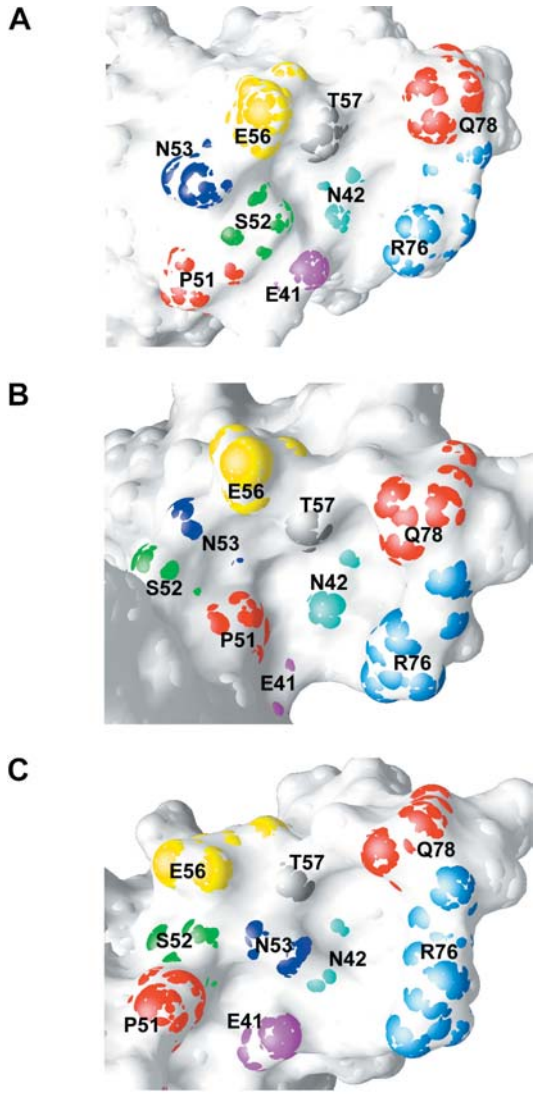

Figure 5. Oligosaccharide-binding pocket architecture, with residues involved in protein-ligand interaction numbered. (*A*) Compact solution-structure monomer (1iiy); (*B*) trigonal crystal structure (3ezm); (*C*) tetragonal crystal structure (1l5b). Figures generated with program GRASP [60].

domain orientation and do not bind oligosaccharide in the primary site [16a]. The high pH (tetragonal) crystal structures have a different relative orientation of the domains but still show a perturbed binding pocket. In this case, the side chain of Asn53 from the hinge moves some way into the pocket, reducing its volume (fig. 5C) and presenting an unfavorable geometry for the specific binding of a mannose ring. Because of these steric considerations, sugar binding in this pocket may not be favorable and was not observed in any of the tetragonal crystal structures at various pH values. Since we do not observe binding into the primary sugar binding site, the question arises whether the binding into this pocket is also carried out by three stacked rings as in the secondary binding site, or by two stacked rings only, as previously reported [11]. There is also the possibility of a different sugar-binding mode for the two sites [16a].

The secondary sugar binding site is located ~35 Å from the primary site located on the same monomer or quasi-monomer. The binding interface is formed by three $\alpha 1 \rightarrow 2$ linked, stacked rings in the case of Man-9 (fig. 3B) and two in the case of hexamannose (fig. 3C) [16a]. This observation is in agreement with other results which show that the oligosaccharides interface with CV-N via a branch containing mainly $\alpha 1 \rightarrow 2$ linked rings. The position of mannose rings C and 4 from arm D1 is very similar to the model suggested from the solution structure of a CV-N disaccharide complex [11]. However, the stacked rings interact much more tightly with the protein than is suggested in the NMR model, with nine hydrogen bonds between the Man-9 rings and CV-N. Since the only difference between the D1 branches in Man-9 and hexamannose is the lack of one $\alpha 1 \rightarrow 2$ linked mannose ring at the end of the branch in the latter, the crystal structures unequivocally identify the D1 arm as the binding interface.

Since the primary sugar binding site is in close proximity to the hinge region, mutations in this region can alter the shape of the binding pocket, influencing sugar binding to CV-N. The secondary sugar binding site, which is located far from the hinge region, is not affected by the geometry of the hinge, presenting the same conformation in both the monomeric and in domain-swapped dimeric CV-N [16a].

Comparing the coordinates of the tetragonal, trigonal and the double-mutant structures, the largest differences are located around the carbohydrate binding sites [16]. Circular dichroism (CD) measurements show that conformational changes occur upon CV-N binding to gp120 and gp41 [7, 31, 32], resulting in an average 11% loss of β sheet and 2% loss of helical structure. In an attempt to determine whether this reflects in the flexibility of the oligosaccharide-binding site in the uncomplexed and sugar-bound forms of CV-N, the B-factors of the respective structures were compared. A well-conserved trend of higher than average B-factors was observed in the regions containing the residues directly involved in creating the sugar-binding sites.

CVN mutants and other constructs

Anti-HIV activity and gp120 binding have been determined not only for the wild-type CV-N, but also for a series of engineered mutants and other constructs [33] (table 2). While the binding of gp120 may be essential, it is not a sufficient requirement for anti-HIV activity of CV-N. The sequence specificity for anti-HIV activity and gp120 binding do not appear to be identical. Some of the early constructs established the key structural regions re-

Table 2. Summary of biological activity of CV-N and its mutants: F indicates uncleaved FLAG-tag used in protein expression [33]; the second column shows the order of residues in the various constructs; the XTT-tetrazolium assay *,†,‡ [61], and an HIV-1 envelope-mediated cell fusion assay § were used for biological activity determination.

Protein	Order of residues	EC_{50} (nM)
wild-type CVN*	1–101	0.9 ± 0.4
CVN-1*	2–101	2.0 ± 0.7
CVN-2*	3–101	8.3 ± 4.0
CVN-3*	4–101	140.7 ± 36.7
F-CVN*	1–101	3.6 ± 1.4
F-CVN*	1–98	149.5 ± 17.2
F-CVN*	1–93	inactive
F-D1D1*	1–50, 1–50	inactive
F-D2D2*	52–101, 51–101	inactive
F-D2D1*	52–101, 1–50	217.7 ± 9.5
F-C8S-C22S*	1–101	inactive
F-C58S-C73S*	1–101	inactive
		EC_{3D} (nM)
wild type CV-N†	1–101	0.19
A77T†	1–101	0.22
S52P†	1–101	1.46
cpCVN‡	1–3, 54–98, 48–53, 4–47, 99–101	1000 × less potent than wt
dsCVN§	1–3, 54–98, 48–53, 4–47, 99–101	500 × less potent than wt

Compiled from *[33], †[34], ‡[39] and §[38]. Since activity values were obtained using different assays, they are not directly comparable.

sponsible for activity [33]. However, structural studies of CV-N opened the possibility of rationally designing specific mutants, which could drastically alter some of the characteristics of the protein. Mutations in the hinge region can influence the dynamics of the two domains, preventing 3D domain swapping or locking the protein in an extended conformation. Since the primary sugar binding site is in close proximity to the hinge region, some of these mutations can alter the shape of the binding pocket, preventing sugar binding to CV-N. Most of these mutations in the hinge region have no effect on the sugar binding to the secondary sugar binding site, which is located farther from the hinge region.

A hinge-region mutant of CV-N, S52P, was discovered in a screen of phage-displayed mutant CV-N library. It has been shown that this mutant exists in solution exclusively as a stable dimer [34]. S52P shows lower potency against HIV compared with the wild-type CV-N. Another engineered mutant, P51G, is mainly monomeric in solution with only trace amounts of dimers and is more stable than wild-type CV-N under certain experimental conditions [5].

A double mutant P51S/S52P of CV-N was engineered by swapping two critical hinge-region residues, Pro51 and Ser52. This mutant has biochemical and biophysical characteristics equivalent to the wild-type CV-N, and its crystal structure resembles that of wild-type CV-N [16]. However, the mutant shows a different orientation in the hinge region that connects two domains of the protein. The observation that this double mutant crystallizes under a wide variety of conditions sheds different light on some of the current hypotheses on domain swapping and on the role of hinge-region proline residues in domain orientation [16].

An obligate domain-swapped mutant was designed by deleting Gln50 from the hinge region [35]. Unlike wild-type CV-N, this mutant is domain-swapped regardless of pH and is dimeric as shown by NMR relaxation experiments. In a quantitative vaccinia virus-based HIV-1 fusion assay this mutant is a slightly more potent inhibitor of HIV-1 fusion than monomeric wild-type CV-N [35]. Relevance of the latter result to the more pertinent question of actual antiviral activity against clinical HIV isolates is unknown.

A series of cumulative mutants of the secondary binding site of CV-N were recently designed, and their overall structure and carbohydrate binding ability were characterized by NMR [36]. Carbohydrate binding in the secondary site was completely abolished in a triple (K3N, E23I, N93A) and a quadruple (K3N, T7A, E23I, N93A) mutant, while Manα1-2Manα was bound in the primary site as strongly as in the wild-type CV-N. All of the mutants inhibited HIV-1 fusion with IC_{50} (inhibitory concentration 50%) values nearly identical to those of wild-type CV-N, as measured in an HIV-1 cell fusion assay [37]. These results were interpreted as indicating that the secondary binding site is not necessary for high-affinity binding to gp120, with monovalent protein-oligosaccharide interactions being sufficient for blocking HIV-1 fusion [36].

An internal sequence-swapped mutant (dsCV-N) in which the cores of domain A and B have been swapped, but the first and the last three amino acids and the linker remain unaltered [38], has been constructed. In a quantitative

HIV-1 envelope-mediated cell fusion assay [37], dsCV-N was 500-fold less active than wild-type CV-N. Using residual dipolar couplings, it was shown that dsCV-N has the same overall shape and possibly structure as wild-type CV-N [38]. This construct can also be described as a variation of a circularly-permuted form of CV-N (cpCV-N) in which several residues are shifted in their sequence [12, 39]. The monomeric cpCV-N was shown to be moderately less stable than the wild-type CV-N and to display significantly reduced anti-HIV activity. Preliminary NMR experiments showed that cpCV-N adopts the same fold as wild-type CV-N, and the determination of a detailed, high-resolution structure is in progress [39].

Antiviral properties of CV-N

The studies of structural and biological properties of CV-N have established that they are all dependent on the unique carbohydrate-binding properties of this protein. CV-N preferentially inhibits binding of the glycosylation-dependent neutralizing monoclonal antibody 2G12 to gp120 [3, 32], and also interferes with the binding of cell-associated CD4 to virion-associated gp120, but not of the soluble CD4 (sCD4) with either soluble gp120 (sgp120) or virion-associated gp120 [40]. CV-N impairs both CD4-dependent and CD4-independent binding of gp120 to the target cells, blocks the sCD4-induced binding of sgp120 with cell-associated coreceptor CXCR4 and dissociates bound sgp120 from target cells. Effects of CV-N on gp120-CD4 binding depend upon the type of CD4 (soluble or cell-associated), but not upon the type of gp120 (soluble or virion associated) [40].

The ability of CV-N to distinguish between high-mannose and complex oligosaccharides may account for the specificity of binding of CV-N to gp120 [41]. The exact binding region on gp120 is yet to be determined. The binding sites for CV-N are not identical, but they might overlap in regions containing high mannose [32].

Association of CV-N with gp120 is necessary but not sufficient for its antiviral activity [33]. Consistent with this observation is the wide range of antiretroviral activity of CV-N, potentially inhibiting all laboratory strains of HIV-1, HIV-2, SIV and primary isolates, which are highly resistant to neutralization by sCD4 or by most anti-gp120 antibodies. There is no direct indication of the lethality of CV-N to the host cells at the concentration of 45–400 nM, with slight effect on viable cell numbers at 9000 nM [3]. The anti-HIV activity of CV-N can be completely abrogated by pretreatment with sgp41 (IC_{50} = 24 μM) and sgp120 (IC_{50} = 0.9 μM) [7]. Isothermal calorimetry and optical biosensor binding studies showed that CV-N bound to recombinant sgp120 with a K_d of 2–45 nM and a 5:1 stoichiometry, and to sgp41 with K_d of 606 nM and 1:1 stoichiometry [7].

Although CV-N is a prokaryotic protein, it efficiently blocks in vitro infection and replication of HIV and other retroviruses that presumably target eukaryotic hosts. It is known that cyanobacteria, including members of the genus *Nostoc,* can be infected by viruses [42–45]. However, the physiological role of CV-N within cyanobacteria and the significance, if any, of CV-N inhibiting retroviruses are presently unknown [9].

Multivalent binding

Titration calorimetric binding experiments of CV-N to Man-8 show an enthalpically driven binding [41]. It was shown that protein lectins and oligosaccharides can overcome weak individual interaction free energies of binding by mediating multivalent binding with one another [46, 47]. Such an effect might explain the unusually tight binding of Man-8 to CV-N, supported by an increase in the molecular ordering and formation of cross-linked precipitates [41]. In addition to the protein-oligosaccharide interactions, specific protein-protein interactions may possibly play an important role in enhancing the CV-N gp120 binding, since sgp120 mediates a 13-fold greater affinity with CV-N than Man-8 [41]. The following CV-

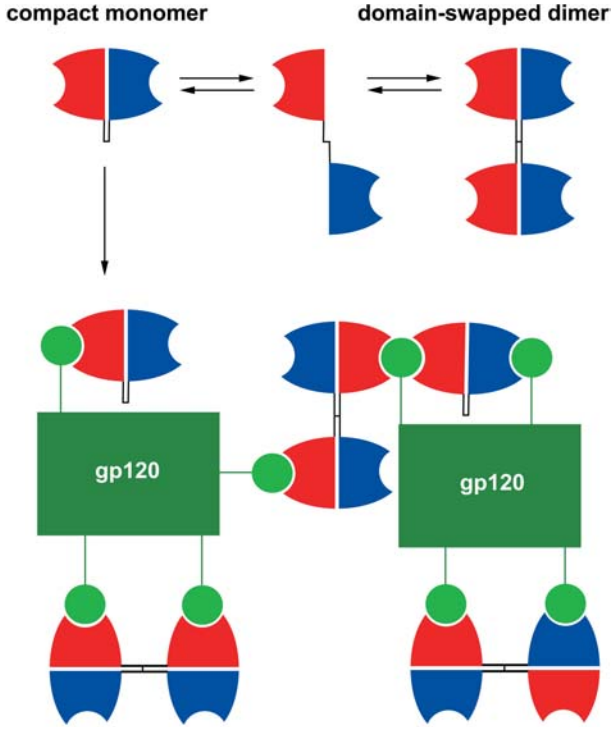

Figure 6. Hypothetical view of possible CV-N gp120 interaction modes. The CV-N domain with the primary site is shown in red, while the one with the secondary site is shown in blue. The compact CV-N monomer can exist in equilibrium with the domain-swapped dimer, and both species can bind to oligomannose on gp120 (green).

N/gp120 binding model was proposed [26]. Given the presence of two oligosaccharide binding sites with different affinities, it is likely that CV-N binds to high-mannose saccharides on gp120 via the high-affinity site. Once bound, due to the high apparent oligomannose concentration, CV-N could further interact with an appropriately spaced second oligomannose via the lower-affinity site. Binding to oligomannose on gp120 through both oligosaccharide binding sites will result in multivalent binding that can facilitate cross-linking, observed with Man-8 and Man-9 at micromolar to millimolar concentrations [26]. The model, together with other possible interaction modes, are illustrated in figure 6. Due to domain swapping and change in the relative quasi-monomer orientation, there are a series of possible CV-N/gp120 interaction modes via the high-affinity sites or the high- and low-affinity sites simultaneously. It is possible that local pH conditions regulate gp120 binding by influencing the quasi-monomer orientation.

Potential clinical applications of CV-N

Due to its broad spectrum of activity against immunodeficiency viruses, CV-N is promising in both prevention and treatment strategies for AIDS. Seventy-five to eighty-five percent of the more than 30 million people infected with HIV were infected by transmission of the virus through heterosexual contact [48]. HIV has a high mutation rate, rapidly evolving drug-resistant strains after administration of most anti-HIV drugs, further complicating development of vaccines [49, 50]. A drug that could effectively prevent initial viral infection would be more likely to prevent the emergence of a drug-resistant virus. There is a growing interest in development of anti-HIV microbicides, that may be used either topically or ex vivo to prevent the spread of viral infection [51], CV-N being one of the prime candidates [3, 6, 32].

In addition to direct virucidal applications, CV-N could be used in creating a recombinant chimeric toxin molecule in which the translocation and cytotoxic domains of *Pseudomonas* exotoxin A are linked to CV-N [52]. In the resulting molecule, CV-N serves as the targeting moiety that binds to HIV-infected cells expressing gp120. The chimeric molecule shows enhanced cytotoxicity for HIV-infected H9 cells compared with uninfected H9 cells, although with diminution of the overall gp120 binding activity. It may be possible to attach low molecular weight, nonproteinaceous cytotoxic agents to CV-N that would not diminish its gp120 binding activity [52].

Since CV-N might be immunogenic in humans, methods were explored for its external use for ex vivo inactivation of HIV in blood or plasma [53]. Biotinylated CV-N was coupled to streptavidin-coated magnetic beads to provide sessile CV-N. Sessile CV-N completely removed 100 $TCID_{50}$ (median tissue culture infectious dose) of the infectious virions but left behind a relatively large fraction of replication-incompetent virions. The latter fraction could be used as a vaccine for HIV. CV-N is currently in preclinical development for use as a topical microbicide to prevent genital HIV transmission [4].

An effective CV-N-based microbicide may require long-term presence or continuous delivery of CV-N at the mucosal sites of entry of HIV. A possible approach is endogenous production of virucidal concentrations of CV-N at mucosal surfaces by a colonizing commensal bacterium such as *Streptococcus gordonii* [54], which is a well-studied vector for mucosal delivery of heterologous proteins [55]. Two recombinant strains were engineered, expressing CV-N as a cell-surface-anchored protein or as a protein secreted into the culture medium. Soluble recombinant CV-N secreted by *S. gordonii* proved to be capable of binding gp120 in a concentration-dependent manner, whereas recombinant bacteria expressing CV-N on their surface were able to capture HIV virions efficiently in an in vitro HIV capturing assay. Work is in progress exploring the possibility of expression of CV-N in *Lactobacillus* strains [54]. It is likely that the extensive knowledge of the structure of CV-N and its mode of interactions with oligosaccharides will be a driving force in its development as a novel antiviral agent.

Acknowledgements. The writing of this review benefited from our past and present collaboration with L. G. Barrientos, C. A. Bewley, M. R. Boyd, G. M. Clore, A. M. Gronenborn, T. Mori and B. R. O'Keefe. We thank M. R. Boyd for critical reading of the manuscript. Work in this laboratory was supported in part by the Intramural AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health.

1 Demain A. L. (1999) Pharmaceutically active secondary metabolites of microorganisms. Appl. Microbiol. Biotechnol. **52:** 455–463
2 O'Keefe B. R. (2001) Biologically active proteins from natural product extracts. J. Nat. Prod. **64:** 1373–1381
3 Boyd M. R., Gustafson K. R., McMahon J. B., Shoemaker R. H., O'Keefe B. R., Mori T. et al. (1997) Discovery of cyanovirin-N, a novel human immunodeficiency virus- inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. Antimicrob. Agents Chemother. **41:** 1521–1530
4 De Clercq E. (2000) Current lead natural products for the chemotherapy of human immunodeficiency virus (HIV) infection. Med. Res. Rev. **20:** 323–349
5 Mori T., Gustafson K. R., Pannell L. K., Shoemaker R. H., Wu L., McMahon J. B. et al. (1998) Recombinant production of cyanovirin-N, a potent human immunodeficiency virus-inactivating protein derived from a cultured cyanobacterium. Protein Expr. Purif. **12:** 151–158
6 Dey B., Lerner D. L., Lusso P., Boyd M. R., Elder J. H. and Berger E. A. (2000) Multiple antiviral activities of cyanovirin-N: blocking of human immunodeficiency virus type 1 gp120 interaction with CD4 and coreceptor and inhibition of diverse enveloped viruses. J. Virol. **74:** 4562–4569
7 O'Keefe B. R., Shenoy S. R., Xie D., Zhang W., Muschik J. M., Currens M. J. et al. (2000) Analysis of the interaction between

the HIV-inactivating protein cyanovirin-N and soluble forms of the envelope glycoproteins gp120 and gp41. Mol. Pharmacol. **58:** 982–992
8 Bergsma J., Hol W. G., Jansonius J. N., Kalk K. H., Ploegman J. H. and Smit J. D. (1975) The double domain structure of rhodanese. J. Mol. Biol. **98:** 637–643
9 Gustafson K. R., Sowder R. C., Henderson L. E., Cardellina J. H., McMahon J. B., Rajamani U. et al. (1997) Isolation, primary sequence determination and disulfide bond structure of cyanovirin-N, an anti-HIV (human immunodeficiency virus) protein from the cyanobacterium *Nostoc ellipsosporum*. Biochem. Biophys. Res. Commun. **238:** 223–228
10 Bewley C. A., Gustafson K. R., Boyd M. R., Covell D. G., Bax A., Clore G. M. et al. (1998) Solution structure of cyanovirin-N, a potent HIV-inactivating protein. Nat. Struct. Biol. **5:** 571–578
11 Bewley C. A. (2001) Solution structure of a cyanovirin-N:Man alpha 1-2Man alpha complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. Structure. (Camb.) **9:** 931–940
12 Barrientos L. G., Campos-Olivas R., Louis J. M., Fiser A., Sali A. and Gronenborn A. M. (2001) 1H, 13C, 15N resonance assignments and fold verification of a circular permuted variant of the potent HIV-inactivating protein cyanovirin-N. J. Biomol. NMR **19:** 289–290
13 Clore G. M. and Bewley C. A. (2002) Using conjoined rigid body/torsion angle simulated annealing to determine the relative orientation of covalently linked protein domains from dipolar couplings. J. Magn. Reson. **154:** 329–335
14 Barrientos L., Louis J. M., Botos I., Mori T., Han Z., O'Keefe B. R. et al. (2002) The domain-swapped dimer of cyanovirin-N is in a metastable folded state. Reconcilliation of X-ray and NMR structures. Structure **10:** 673–686
15 Yang F., Bewley C. A., Louis J. M., Gustafson K. R., Boyd M. R., Gronenborn A. M. et al. (1999) Crystal structue of cyanovirin-N, a potent HIV-inactivating protein, shows unexpected domain swapping. J. Mol. Biol. **288:** 403–412
16 Botos I., Mori T., Cartner L. K., Boyd M. R. and Wlodawer A. (2002) Domain-swapped structure of a mutant of cyanovirin-N. Biochem. Biophys. Res. Commun. **294:** 184–190
16a Botos I., O'Keefe B. R., Shenoy S. R., Cartner L. K., Ratner D. M., Seeberger P. H. et al. (2002) Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high-mannose oligosaccharides. J. Biol. Chem. **277:** 1–7
17 Clore G. M. and Gronenborn A. M. (1998) New methods of structure refinement for macromolecular structure determination by NMR. Proc. Natl. Acad. Sci. USA **95:** 5891–5898
18 Gooptu B., Hazes B., Chang W. S., Dafforn T. R., Carrell R. W., Read R. J. et al. (2000) Inactive conformation of the serpin alpha(1)-antichymotrypsin indicates two-stage insertion of the reactive loop: implications for inhibitory function and conformational disease. Proc. Natl. Acad. Sci. USA **97:** 67–72
19 Bewley C. A. and Clore G. M. (2000) Determination of the relative orientation of the two halves of the domain-swaped dimer of cyanovirin-N in solution using dipolar couplings and rigid body minimization. J. Am. Chem. Soc. **122:** 6009–6016
20 Holm L. and Sander C. (1993) Protein structure comparison by alignment of distance matrices. J. Mol. Biol. **233:** 123–138
21 Edmondson S. P., Qiu L. and Shriver J. W. (1995) Solution structure of the DNA-binding protein Sac7d from the hyperthermophile *Sulfolobus acidocaldarius*. Biochemistry **34:** 13289–13304
22 Kuriyan J. and Cowburn D. (1993) Structures of SH2 and SH3 domains. Curr. Opin. Struct. Biol. **3:** 828–837
23 Bennett M. J., Schlunegger M. P. and Eisenberg D. (1995) 3D domain swapping: a mechanism for oligomer assembly. Protein Sci. **4:** 2455–2468
24 MacArthur M. W. and Thornton J. M. (1991) Influence of proline residues on protein conformation. J. Mol. Biol. **218:** 397–412
25 Rousseau F., Schymkowitz J. W., Wilkinson H. R. and Itzhaki L. S. (2001) Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. Proc. Natl. Acad. Sci. USA **98:** 5596–5601
26 Bewley C. A. and Otero-Quintero S. (2001) The potent anti-HIV protein cyanovirin-N contains two novel carbohydrate binding sites that selectively bind to Man(8) D1D3 and Man(9) with nanomolar affinity: implications for binding to the HIV envelope protein gp120. J. Am. Chem. Soc. **123:** 3892–3902
27 Bolmstedt A. J., O'Keefe B. R., Shenoy S. R., McMahon J. B. and Boyd M. R. (2001) Cyanovirin-N defines a new class of antiviral agent targeting N-linked, high-mannose glycans in an oligosaccharide-specific manner. Mol. Pharmacol. **59:** 949–954
28 Hadden J. M., Convery M. A., Declais A. C., Lilley D. M. and Phillips S. E. (2001) Crystal structure of the Holliday junction resolving enzyme T7 endonuclease I. Nat. Struct. Biol. **8:** 62–67
29 Declais A. C., Hadden J., Phillips S. E. and Lilley D. M. (2001) The active site of the junction-resolving enzyme T7 endonuclease I. J. Mol. Biol. **307:** 1145–1158
30 Chen Y., Narendra U., Iype L. E., Cox M. M. and Rice P. A. (2000) Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. Mol. Cell **6:** 885–897
31 Mariner J. M., McMahon J. B., O'Keefe B. R., Nagashima K. and Boyd M. R. (1998) The HIV-inactivating protein, cyanovirin-N, does not block gp120- mediated virus-to-cell binding. Biochem. Biophys. Res. Commun. **248:** 841–845
32 Esser M. T., Mori T., Mondor I., Sattentau Q. J., Dey B., Berger E. A. et al. (1999) Cyanovirin-N binds to gp120 to interfere with CD4-dependent human immunodeficiency virus type 1 virion binding, fusion, and infectivity but does not affect the CD4 binding site on gp120 or soluble CD4- induced conformational changes in gp120. J. Virol. **73:** 4360–4371
33 Mori T., Shoemaker R. H., Gulakowski R. J., Krepps B. L., McMahon J. B., Gustafson K. R. et al. (1997) Analysis of sequence requirements for biological activity of cyanovirin-N, a potent HIV (human immunodeficiency virus)- inactivating protein. Biochem. Biophys. Res. Commun. **238:** 218–222
34 Han Z., Xiong C., Mori T. and Boyd M. R. (2002) Discovery of a stable dimeric mutant of cyanovirin-N (CV-N) from a T7 phage-displayed CV-N mutant library. Biochem. Biophys. Res. Commun. **292:** 1036–1043
35 Kelley B. S., Chang L. C. and Bewley C. A. (2002) Engineering an obligate domain-swapped dimer of cyanovirin-N with enhanced anti-HIV activity. J. Am. Chem. Soc. **124:** 3210–3211
36 Chang L. C. and Bewley C. A. (2002) Potent inhibition of HIV-1 fusion by cyanovirin-N requires only a single high affinity carbohydrate binding site: characterization of low-affinity carbohydrate binding site knockout mutants. J. Mol. Biol. **318:** 1– 8
37 Nussbaum O., Broder C. C. and Berger E. A. (1994) Fusogenic mechanisms of enveloped-virus glycoproteins analyzed by a novel recombinant vaccinia virus-based assay quantitating cell fusion-dependent reporter gene activation. J. Virol. **68:** 5411–5422
38 Bewley C. A. (2001) Rapid validation of the overall structure of an internal domain-swapped mutant of the anti-HIV protein cyanovirin-N using residual dipolar couplings. J. Am. Chem. Soc. **123:** 1014–1015
39 Barrientos L. G., Louis J. M., Hung J., Smith T. H., O'Keefe B. R., Gardella R. S. et al. (2002) Design and initial characterization of a circular permuted variant of the potent HIV-inactivating protein cyanovirin-N. Proteins **46:** 153–160
40 Mori T. and Boyd M. R. (2001) Cyanovirin-N, a potent human immunodeficiency virus-inactivating protein, blocks both CD4-dependent and CD4-independent binding of soluble gp120 (sgp120) to target cells, inhibits sCD4-induced binding

of sgp120 to cell-associated CXCR4, and dissociates bound sgp120 from target cells. Antimicrob. Agents Chemother. **45:** 664–672
41 Shenoy S. R., O'Keefe B. R., Bolmstedt A. J., Cartner L. K. and Boyd M. R. (2001) Selective interactions of the human immunodeficiency virus-inactivating protein cyanovirin-N with high-mannose oligosaccharides on gp120 and other glycoproteins. J. Pharmacol. Exp. Ther. **297:** 704–710
42 Hu N. T., Thiel T., Giddings T. H. Jr and Wolk C. P. (1981) New *Anabaena* and *Nostoc* cyanophages from sewage settling ponds. Virology **114:** 236–246
43 Sherman L. A. and Connelly M. (1976) Isolation and characterization of a cyanophage infecting the unicellular blue-green algae *A. nidulans* and *S. cedrorum*. Virology **72:** 540–544
44 Adolph K. W. and Haselkorn R. (1972) Photosynthesis and the development of blue-green algal virus N-1. Virology **47:** 370–374
45 Safferman R. S., Morris M. E., Sherman L. A. and Haselkorn R. (1969) Serological and electron microscopic characterization of a new group of blue-green algal viruses (LPP-2). Virol. **39:** 775–780
46 Rice K. G. and Lee Y. C. (1993) Oligosaccharide valency and conformation in determining binding to the asialoglycoprotein receptor of rat hepatocytes. Adv. Enzymol. Relat Areas Mol. Biol. **66:** 41–83
47 Dimick S. M., Powel S. C., McMahon S. A., Moothoo D. N., Haismith J. H. and Toone E. J. (2002) On the meaning of affinity: cluster glycoside effects and concanavalin A. J. Am. Chem. Soc **121:** 10286–10296
48 Piot P. (1998) The science of AIDS: a tale of two worlds. Science **280:** 1844–1845
49 Burton D. R. (1997) A vaccine for HIV type 1: the antibody perspective. Proc. Natl. Acad. Sci. USA **94:** 10018–10023
50 Heilman C. A. and Baltimore D. (1998) HIV vaccines – where are we going? Nat. Med. **4:** 532–534
51 Lange J. M., Karam M. and Piot P. (1993) Boost for vaginal microbicides against HIV. Lancet **342:** 1356
52 Mori T., Shoemaker R. H., McMahon J. B., Gulakowski R. J., Gustafson K. R. and Boyd M. R. (1997) Construction and enhanced cytotoxicity of a [cyanovirin-N]- [*Pseudomonas* exotoxin] conjugate against human immunodeficiency virus-infected cells. Biochem. Biophys. Res. Commun. **239:** 884–888
53 Gandhi M. J., Boyd M. R., Yi L., Yang G. G. and Vyas G. N. (2000) Properties of cyanovirin-N (CV-N): inactivation of HIV-1 by sessile cyanovirin-N (sCV-N). Dev. Biol. Stand. **102:** 141–148
54 Giomarelli B., Provvedi R., Meacci F., Maggi T., Medaglini D., Mori T. et al. (2002) The microbicide cyanovirin-N expressed in the surface of commensal bacterium *Streptococcus gordonii* captures HIV-1. AIDS **16:** 1351–1356
55 Oggioni M. R., Manganelli R., Contorni M., Tommasino M. and Pozzi G. (1995) Immunization of mice by oral colonization with live recombinant commensal streptococci. Vaccine **13:** 775–779
56 Esnouf R. M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. J. Mol. Graph. Model. **15:** 132–134
57 Meritt E. A. and Murphy M. E. P. (1994) Raster3D Version 2.0. A program for photorealistic molecular graphics. Acta. Crystallogr. **D50:** 869–873
58 Cohen G. E. (1997) ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. J. Appl. Crystallogr. **30:** 1160–1161
59 Christopher, J. A. (1998) SPOCK: The Structural Properties Observation and Calculation Kit. The Center for Macromolecular Design, Texas A&M University, College Station, TX
60 Nicholls, A. (1992) GRASP: Graphical Representation and Analysis of Surface Properties. Columbia University, New York
61 Gulakowski R. J., McMahon J. B., Staley P. G., Moran R. A. and Boyd M. R. (1991) A semiautomated multiparameter approach for anti-HIV drug screening. J. Virol. Methods **33:** 87–100